Changes to the End-Stage Renal Disease Quality Incentive Program

Steven Fishbane[1], Ilene Miller[1], John D. Wagner[1] and Naveed N. Masani[2]

[1]North Shore-LIJ Health System, Department of Medicine, Division of Nephrology, Hofstra North Shore-LIJ School of Medicine, Great Neck, New York, USA and [2]Division of Nephrology, Winthrop-University Hospital, Mineola, New York, USA

Monitoring the quality of dialysis care has long been a component of the Medicare ESRD program. As part of the 2008 Medicare Improvements for Patients and Providers Act (MIPPA), Congress mandated the Quality Incentive Program (QIP), which linked measures of care quality to payments. The legislation embraced the idea that this linkage of federal money to performance would encourage the purchase of greater 'value.' The first 2 program years for the QIP use a simple scoring methodology and a limited scope of quality metrics. For payment year 2014 (performance period calendar year 2012), the program changes substantially, with an expanded number of quality measures and a more complex scoring methodology. In this article, we describe the program structure, quality measures, scoring system, and financial impact.

Kidney International (2012) **81,** 1167–1171; doi:10.1038/ki.2012.138; published online 25 April 2012

KEYWORDS: anemia; dialysis; quality

Correspondence: *Steven Fishbane, North Shore-LIJ Health System, Department of Medicine, Division of Nephrology, Hofstra North Shore-LIJ School of Medicine,100 Community Drive, Great Neck, New York 11021, USA. E-mail: sfishbane@gmail.com or sfishbane@metrorenal.com*

Received 23 December 2011; revised 8 February 2012; accepted 6 March 2012; published online 25 April 2012

BACKGROUND

The Medicare and Medicaid Programs, managed through the Centers for Medicare and Medicaid Services (CMS), has been consistently clear in its intent to increasingly move from a passive payer for health care to a partner that works with providers to ensure the quality of care received by program beneficiaries.[1] The CMS seeks to attain the triple aim of better health care for individuals, better care for populations and communities, and lower costs[2] through care improvement. In the End-Stage Renal Disease (ESRD) realm, the involvement of CMS in this regard has been long-standing. In recent years, the relationship has become progressively more engaged, first with the public reporting of ESRD quality data[3] and, more recently, with the ESRD Quality Incentive Program (QIP). This program is a requirement of section 1881(h) of the Social Security Act and, remarkably, is one of the first national pay for performance programs of any type.

For payment years 2012 and 2013 (performance years 2010 and 2011), the ESRD QIP is quite simple, based only on anemia and dialysis adequacy measures. For the payment year 2013 (performance 2011), quality is measured simply as the percentage of patients with a urea reduction ratio (URR) of $\geqslant$65% and those with a hemoglobin (Hgb) level of $>$12 g/dl. Payments can be reduced up to a maximum of 2% for failure to achieve acceptable results compared with either national averages or the program's own previous performance.

PAYMENT YEAR 2014 (PERFORMANCE PERIOD 2012)

The ESRD QIP changes significantly in payment year 2014. The two major areas of change are (1) a larger spectrum and number of quality measures and (2) a more advanced, and complex, scoring methodology. We will begin by discussing individual measures with a focus on the scientific merits of each indicator. There are three clinical quality measures, which together account for 90% of the performance score, and three reporting measures, which comprise the remaining 10% of the performance score.

CLINICAL MEASURES

(1) Percentage of erythropoietin-stimulating agent (ESA)—treated patients with Hgb $>$12 g/dl. The higher the percentage, the lower the score for this measure (see

Figure 1 for national changes in Hgb concentrations over time). There are key exclusions, for which we refer the reader to the final program rule.[1]

(2) Percentage of patients with the URR of $\geqslant$65%. The CMS had proposed changing the dialysis adequacy measure to KT/V, but in the final rule reverted to the URR measure owing to lack of standardization for KT/V.

(3) Vascular access type. This measure comprises the percentage of hemodialysis patients using an arteriovenous fistula (AVF) with two needles during the last treatment of the month and the percentage of hemodialysis patients with an intravenous catheter in use for 90 days before the last dialysis session, with no AVF or arteriovenous graft (AVG). Although access type at initiation of dialysis is not counted in the measure, because most US hemodialysis patients initiate treatment via a catheter,[4] success with this measure requires timely removal of the catheter (Figure 2).

The CMS had proposed adding two additional clinical measures, vascular access infections and the standardized hospitalization ratio, but chose in the final rule to postpone inclusion of both indicators.

REPORTING MEASURES

(1) The first reporting measure is based on the dialysis facility's reporting dialysis safety events to the CDC National Healthcare Safety Network.

(2) The second is the attestation that patient satisfaction is assessed by the In-Center Hemodialysis (ICH) Consumer Assessment of Healthcare Providers and Systems (CAHPS) survey.

(3) The final measure is the attestation that serum calcium and phosphorus are measured on at least a monthly basis.

SCIENTIFIC BASIS FOR THE MEASURES

Can the measurement of performance improve medical quality of care? To achieve this goal, Chassin *et al.*[5] have suggested four key principles derived from previous studies that proposed that quality measures should be vetted against:

(1) '… a measure must be based on a strong foundation of research showing that the process addressed by the measure, when performed correctly, leads to improved clinical outcomes.'

(2) 'the measurement strategy must accurately capture whether the evidence-based care has been delivered.'

(3) 'the measure should address a process quite proximate to the desired outcome, with relatively few intervening processes.'

(4) 'the measure should have minimal or no unintended adverse consequences.'

The QIP clinical quality measures should be tested against these criteria.

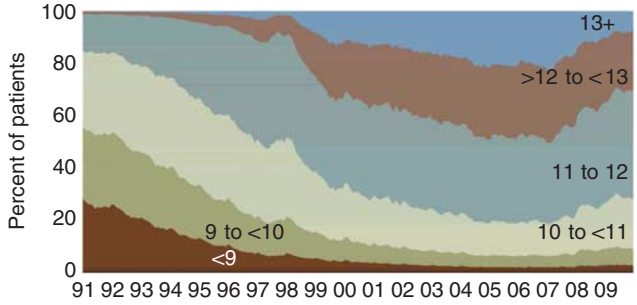

Figure 1 | From the United States Renal Data System (USRDS) 2011 Annual Data Report. The distribution of patients with hemoglobin (Hgb) level >12 g/dl decreased substantially from 2006 to 2009.

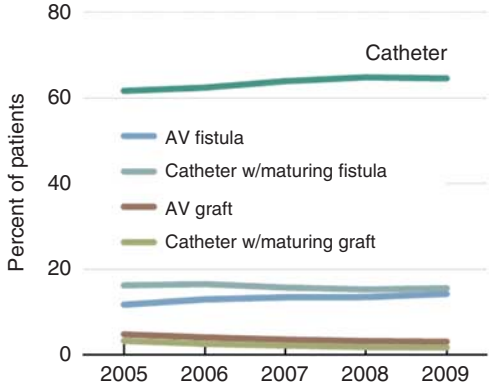

Figure 2 | From the United States Renal Data System (USRDS) 2011 Annual Data Report. At the initiation of hemodialysis, the majority of patients use a dialysis catheter as access at their first dialysis treatment. AV, arteriovenous.

(1) Percentage of patients with Hgb >12 g/dl (the higher the percentage, the lower the score for this measure). The evidence to support this measure comes from a series of randomized controlled trials comparing lower with near-normal Hgb targets. These studies have consistently found adverse outcomes, predominantly cardiovascular and atherothrombotic events when a Hgb target of >13 g/dl is maintained for an extended period, generally several years.[6–8] It is important to note that the CMS hemoglobin quality measure assesses potentially transient increases in Hgb, and not extended targeting. According to the methodology used for the measure, the actual Hgb value used in the determination of whether a patient exceeded 12 g/dl or not is the mean value of all eligible claims (minimum 4 claims) submitted during the year. Therefore, in its most literal interpretation, the Hgb measure violates Chassin principle 1. No RCT (randomized clinical trial) has specifically sought to answer the question of whether transient increases in Hgb to >12 g/dl (or any Hgb level) are harmful to patients. There are some limited observational data,[9] but it can only be concluded that there is no definitive answer to this question. We would suggest that the FDA label, which requires reduction or holding of ESA doses

when Hgb approaches 11 g/dl, is sufficient for addressing potential risk.[10] Indeed, Figure 1 shows that clinical practice has already changed in the direction intended by the FDA. As such, the measure focusing on elevated hemoglobin only provides indirect information regarding prescribing patterns (Chassin principle 2), does not reflect that intercurrent illness, iron deficiency, and/or repletion can alter the relationship between ESA dose and hemoglobin (principle 3), and could have adverse consequences (principle 4) if it is established that the necessarily lower hemoglobins that are incentivized by this scheme do not in fact yield better outcomes.

(2) The percentage of patients with the URR of $\geqslant$65% measures dialysis adequacy. The scientific evidence would seem sufficient to support this indicator; the Kidney Disease Outcomes Quality Initiative guidelines state that the minimally acceptable URR is 65%, with strength of evidence assessed as A.[11] The measure would appear to meet the other Chassin criteria as well, making this an appropriate quality indicator. One concern is the problem of noncompliant patients who sign off treatments early. Although it could be argued that part of the facility's responsibility for quality is to ensure compliance, the extent to which noncompliance with treatment length is fully manageable is unclear. As expressed in terms of quality measures principle 3, an inadequate URR does not always reflect circumstances under facility control, and the quality targets imposed on a facility ideally should accommodate that reality.

(3) The vascular access type clinical indicator averages two subindicators, the use of AV fistulas and the use of dialysis catheters, both measured in patient-months vs. exposure. The individual indicators are both endorsed by the National Quality Forum (NQF). The scientific evidence here is clearly strong.[11,12] One concern could be a violation of Chassin criterion 4, 'the measure should have minimal or no unintended adverse consequences.' A quality indicator that promotes fistulas and penalizes catheters seems reasonable. However, by not including AV grafts as a measure, this indicator could result in attempts to place AV fistulas in inappropriate candidates, which could result in unnecessary discomfort, complications, and increased costs. Neither the vascular access type indicator nor any of the other measures should deter individualization of therapy when appropriate, and this may be accommodated by selecting appropriate benchmarks for performance comparison or permitting clearly defined adjustments for outliers.

Previously, there had been another quality measure—the percentage of patients with Hgb $<$10 g/dl. This was removed with the 2013 payment year. There is a fair body of evidence to suggest that avoidance of Hgb concentrations below 10 g/dl could improve patient outcomes. Among published RCTs, it has been demonstrated that transfusions are reduced[13] and that quality of life probably improves[14,15] with Hgb levels above 10 g/dl (the FDA did not permit such a QOL statement in the ESA labeling, however). This needs to be balanced against a consistent finding of increased cardiovascular risk when the Hgb target is maintained at $>$13 g/dl.[6–8] If one is satisfied that the safety risks when targeting Hgb levels $>$13 g/dl do not extend significantly below 12–13 g/dl, then the balance of risk and benefit supports maintaining the Hgb level above 10 g/dl. This would suggest that a clinical quality indicator measuring the percentage of patients with an Hgb level $>$10 g/dl should be reinstated.

The QIP reporting measures are generally simpler and somewhat exploratory, and contribute only 10% to the Total Performance Score (TPS). Although Section 1881(h)(2)(B)(i) of the Social Security Act generally requires a contracted consensus-based entity's endorsement of measures (a role played by the National Quality Forum (NQF, see below)), it allows certain exceptions 'in the case of a specified area or medical topic determined appropriate by the Secretary for which a feasible and practical measure has not been endorsed by the entity with a contract under section 1890(a).' The reporting measures may be most relevant as an indication of future QIP full weight measures. Caution is certainly warranted here as none of these indicators could be reformulated in an evidence-based manner. For future years, the CMS would do well to eliminate reporting measures and embrace measures that reflect quality in a more direct manner.

Although the Chassin criteria serve as an excellent construct for objective determination of the value of quality measures, NQF endorsement is the statutory requirement for measure inclusion in the QIP. For this reason, it is valuable to understand the NQF methodology. The NQF is a nonprofit organization with widespread membership among healthcare providers and payers. The Chassin criteria are not significantly different from the NQF process for measure endorsement. The NQF also uses four criteria: (1) measures should be in high-priority areas, (2) scientific acceptability, (3) usability and relevance, and (4) feasibility of collection. The NQF does not explicitly characterize the linkage between a quality measure and the improvement process and how confounding factors are to be mitigated.

QUALITY INCENTIVE PROGRAM SCORING SYSTEM

A TPS (range of 0–100) is generated from the total of the individual clinical indicators (90%) and reporting measures (10%). The scoring of clinical indicators is more complex and will be discussed first (Figure 3).

Each of the three clinical indicators receives an achievement score (range of 0–10) and an improvement score (range of 0–9). The higher of the two becomes the performance score for that indicator. The achievement score is calculated based on the facility's performance in 2012. This is compared with national performance during the baseline period, 1 July 2010–30 June 2011. The range of 0–10 is created by setting a 'floor,' the achievement threshold, at the 15th percentile of national performance. Facility performance below this level would result in zero achievement points earned. To achieve

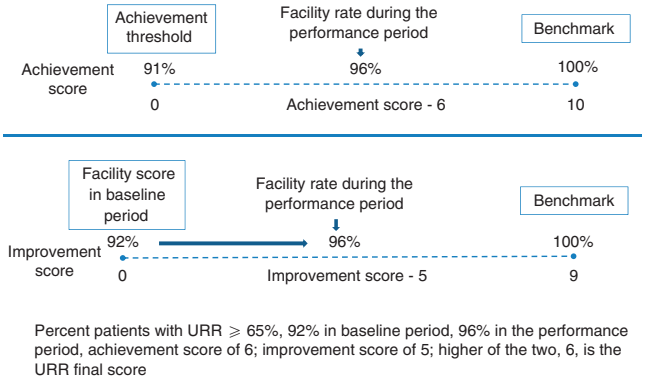

Figure 3 | Dialysis facility percent patients with urea reduction ratio of (URR) ⩾65%, 92% in baseline period, 96% in the performance period, achievement score of 6, improvement score of 5; higher of the two, 6, is the URR final performance score.

the maximal 10 points, the facility would need to perform at or above the national 90th percentile during the baseline period, termed the benchmark. Facility performance between the achievement threshold (0 points) and benchmark (10 points) results in an intermediate score. For example, it is estimated that the achievement threshold for the percentage of patients with a URR ⩾65% will be 91%, and that the benchmark will be 100%. This highlights an important problem with the program's selection of metrics, i.e., the URR measure is 'topped out' by this methodology. If the achievement score is already generally excellent, why select this metric? Its existence seems to serve the purpose of affording an opportunity for CMS to pay less instead of incentivizing good performance. This is because performance below the 100% level will lose points and result in a possible financial penalty. With the achievement threshold for URR ⩾65% at 91% and the benchmark of 100%, a facility with 88% of its patients above a URR of 65% would receive zero points for achievement. If 100% of its patients were above 65% then the score would be 10, and if 96% of its patients were above 65% then the score would be ~6.

The facility improvement score (range of 0–9) for each clinical indicator is intended to give credit for progress made on quality performance. To continue the example from above, if the facility described had 96% of its patients with URR ⩾65% during the performance period (2012) and 92% during the baseline period, and the national benchmark was 100%, then the improvement score would be 5. Because the higher of the achievement or improvement scores becomes the performance score for that clinical indicator, the dialysis adequacy performance score would be 6.

The reporting measures account for 10% of the TPS.

(1) Reporting of safety events (infections) to the CDC's National Healthcare Safety Network results in 5 points for simply enrolling and training in the program, and the maximum 10 points for 3 months of reporting.
(2) Attestation that patient satisfaction is being measured with the ICH CAHPS tool earns 10 points.
(3) Attestation that calcium and phosphorus are tested at least monthly earns 10 points.

TOTAL PERFORMANCE SCORE AND PAYMENT

The TPS determines the amount of payment reduction. It comprises 90% of the clinical indicator scores and 10% of the reporting scores. Each of the three clinical scores (0–10) multiplied by 3 yields a summed clinical indicator score of 0–90. Similarly, the three reporting scores multiplied by 0.33 will add to a sum of 0–10. Together, the TPS can therefore have a final score of 0–100.

Payment can be reduced from 0 to 2% based on the TPS. The CMS has decided that it will base payment on a national minimum TPS. The current estimate is that the minimum TPS will be set at or about 56. Every 10 points the facility TPS scoring falls below this level will result in another 0.5% payment reduction (Table 1).

The financial impact can be approximated by considering total facility Medicare revenues. If a facility has 100 patients and performs 15,600 eligible Medicare treatments per year, and the average payment per treatment is $235, then annual revenues would be $3,666,000. If the TPS is below 26, then 2% of payments, or $73,320, would be lost. We would argue that this financial risk is too small to have a meaningful impact on quality. The Medicare Hospital Inpatient Value Based Purchasing Program (VBP)[16] also has a capped payment risk of 2% (reached in 2017). This program, however, has an interesting financial twist. The 2% is withheld from payments, and then paid back later based on performance. More importantly, it is a zero sum game—that is to say that the amount paid back to the hospital could be substantially more than the 2% initially withheld. It is likely that the VBP program could have a greater impact on quality. On the other hand, the results of the QIP will be public information, and in this regard facilities may be motivated to achieve the performance inherent in the metrics in order to buttress their reputation as providers of quality care.

In conclusion, the ESRD Quality Incentive Program is intended to incent consistently high-quality care for dialysis patients. The 2014 payment year program reflects substantial changes in methodology. Going forward, we would suggest that to the greatest extent possible all future quality measures should be NQF certified and tested against the Chassin criteria. Finally, we would recommend that the system should not simply be punitive, but, similar to the Hospital

Table 1 | Payment year 2014 payment reductions

Total performance score (TPS)	Percentage payment reduction
53–100	No reduction
43–52	0.50%
33–42	1.00%
23–32	1.50%
<23	2.00%

VBP program, should provide the opportunity to increase payment for excellent performance.

DISCLOSURE

All the authors declared no competing interests.

REFERENCES

1. http://www.gpo.gov/fdsys/pkg/FR-2011-11-10/pdf/2011-28606.pdf, accessed 19 December 2011.
2. http://www.ihi.org/offerings/Initiatives/TripleAim/Pages/default.aspx, accessed 22 November 2011.
3. http://www.medicare.gov/Dialysis/Include/DataSection/Questions/SearchCriteria.asp?version=default&browser=IE%7C7%7CWinXP&language=English&defaultstatus=0&pagelist=Home, accessed 23 November 2011.
4. http://www.usrds.org/2011/pdf/v2_ch02_11.pdf, accessed 23 November 2011.
5. Chassin MR, Loeb JM, Schmaltz SP *et al*. Accountability measures–using measurement to promote quality improvement. *N Engl J Med* 2010; **363**: 683–688.
6. Besarab A, Bolton WK, Browne JK *et al*. The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *N Engl J Med* 1998; **339**: 584–590.
7. Singh AK, Szczech L, Tang KL *et al*. Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl J Med* 2006; **355**: 2085–2098.
8. Pfeffer MA, Burdmann EA, Chen CY *et al*., TREAT Investigators. A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *N Engl J Med* 2009; **361**: 2019–2032.
9. Regidor DL, Kopple JD, Kovesdy CP *et al*. Associations between changes in hemoglobin and administered erythropoiesis-stimulating agent and survival in hemodialysis patients. *J Am Soc Nephrol* 2006; **17**: 1181–1191.
10. http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm109375.htm, accessed 23 November 2011.
11. NKF-K/DOQI clinical practice guidelines for hemodialysis adequacy: update 2000. *Am J Kidney Dis* 2001; **37**(Suppl 1): S7–S64.
12. Navuluri R, Regalado S. The KDOQI 2006 vascular access update and fistula first program synopsis. *Semin Intervent Radiol* 2009; **26**: 122–124.
13. Lawler EV, Bradbury BD, Fonda JR *et al*. Transfusion burden among patients with chronic kidney disease and anemia. *Clin J Am Soc Nephrol* 2010; **5**: 667–672.
14. Johansen KL, Finkelstein FO, Revicki DA *et al*. Systematic review and meta-analysis of exercise tolerance and physical functioning in dialysis patients treated with erythropoiesis-stimulating agents. *Am J Kidney Dis* 2010; **55**: 535–548.
15. Keown PA, Churchill DN, Poulin-Costello M *et al*. Dialysis patients treated with epoetin alfa show improved anemia symptoms: a new analysis of the Canadian Erythropoietin Study Group trial. *Hemodial Int* 2010; **14**: 168–173.
16. http://www.cms.gov/Hospital-Value-Based-Purchasing/, accessed 22 November 2011.